Article No. jmbi.1998.2332 available online at http://www.idealibrary.com on IDEAL J. Mol. Biol. (1999) 285, 655–673

JMB

Crystal Structure of dUTPase from Equine Infectious Anaemia Virus; Active Site Metal Binding in a Substrate Analogue Complex

Zbigniew Dauter[1], Rebecca Persson[2], Anna Maria Rosengren[2], Per Olof Nyman[2], Keith S. Wilson[1] and Eila S. Cedergren-Zeppezauer[2]*

[1]*Department of Chemistry University of York, Heslington York, YO1 5DD, UK*

[2]*Biochemistry, Center for Chemistry and Chemical Engineering, Lund University P.O. Box 124, S-221 00 Lund Sweden*

**Corresponding author*

The X-ray structures of dUTPase from equine infectious anaemia virus (EIAV) in unliganded and complexed forms have been determined to 1.9 and 2.0 Å resolution, respectively. The structures were solved by molecular replacement using *Escherichia coli* dUTPase as search model. The exploitation of a relatively novel refinement approach for the initial model, combining maximum likelihood refinement with stereochemically unrestrained updating of the model, proved to be of crucial importance and should be of general relevance.

EIAV dUTPase is a homotrimer where each subunit folds into a twisted antiparallel β-barrel with the N and C-terminal portions interacting with adjacent subunits. The C-terminal 14 and 17 amino acid residues are disordered in the crystal structure of the unliganded and complexed enzyme, respectively. Interactions along the 3-fold axis include a water-containing volume (size 207 $Å^3$) which has no contact with bulk solvent.

It has earlier been shown that a divalent metal ion is essential for catalysis. For the first time, a putative binding site for such a metal ion, in this case Sr^{2+}, is established. The positions of the inhibitor (the non-hydrolysable substrate analogue dUDP) and the metal ion in the complex are consistent with the location of the active centre established for trimeric dUTPase structures, in which subunit interfaces form three surface clefts lined with evolutionary conserved residues. However, a detailed comparison of the active sites of the EIAV and *E. coli* enzymes reveals some structural differences. The viral enzyme undergoes a small conformational change in the uracil-binding β-hairpin structure upon dUDP binding not observed in the other known dUTPase structures.

Keywords: dUTPase; metal-inhibitor complex; equine infectious anaemia virus; X-ray structure; automated refinement procedure

Introduction

The inability of DNA polymerase to discriminate between dUTP and dTTP during DNA synthesis (Bessman *et al.*,1958; Mosbaugh, 1988; Focher *et al.*, 1990; Kornberg & Baker, 1991) and the spontaneous deamination of cytosine present in DNA (Lindahl, 1982) promote the occurrence of the RNA base uracil in DNA. Uracil incorporated into DNA is excised and replaced with the correct base, as encoded by the opposite strand (Tye *et al.*, 1977; Tye & Lehman, 1977; Hochhauser & Weiss, 1978; Sancar & Sancar, 1988). A high concentration of dUTP leads to excessive incorporation of uracil, and the excision repair pathway will expedite a deleterious cycle of multiple excisions, generating mutations and/or strand fragmentation with the possibility of concomitant cell death (Barclay *et al.*, 1982; Richards *et al.*, 1984; Ingraham *et al.*, 1986). Therefore, the maintenance of a low dUTP/dTTP ratio is vital (Goulian *et al.*, 1986).

Abbreviations used: EIAV, equine infectious anaemia virus; FIV, feline immunodeficiency virus; dUTPase, deoxyuridine 5′-triphosphate nucleotidohydrolase; (u)ARP, (unrestrained) automatic refinement procedure; rms, root mean square; 3D, three-dimensional; MMTV, mouse mammary tumour virus.

E-mail address of the corresponding author: eila.cedergren@biokem.w.se

0022-2836/99/020655–19 $30.00/0

A crucial role in this regulation is played by the enzyme dUTP nucleotidohydrolase (dUTPase, EC 3.6.1.23), which catalyses the hydrolysis of dUTP to dUMP and PP_i (Bertani *et al*., 1961, 1967; Greenberg & Somerville, 1962; Shlomai & Kornberg, 1978). The enzyme reduces uracil misincorporation in two ways. Firstly, by lowering the dUTP concentration, it minimises the usage of dUTP in DNA synthesis and repair. Secondly, by producing dUMP, it provides the substrate for the *de novo* biosynthesis of dTTP, further shifting the dUTP:dTTP ratio in a favourable direction.

dUTPase activity has been demonstrated in a wide variety of organisms, from prokaryotes to eukaryotes, and is essential for the viability of *Escherichia coli* (El-Hajj *et al*., 1988) and *Saccharomyces cerevisiae* (Gadsden *et al*., 1993). The expression of cellular dUTPases has been suggested to be cell cycle dependent or developmentally regulated. Actively dividing cells exhibit relatively high levels of endogenous dUTPase activity (Pardo & Gutierrez, 1990; Hokari *et al*., 1982; Strahler *et al*., 1993), while terminally differentiated and/or non-dividing cells contain lower levels of the enzyme (Spector & Boose, 1983; Pardo & Gutierrez, 1990; Pri-Hadash *et al*., 1992; Strahler *et al*., 1993). Several viruses, including pox and herpes and a subset of retroviruses, have been shown to encode their own dUTPases, despite the presence of a host enzyme (Wohlrab & Francke, 1980; Caradonna & Adamkiewicz, 1984; Preston & Fisher, 1984; Williams *et al*., 1985; Mercer *et al*., 1989; Elder *et al*., 1992; Broyles, 1993; Roseman *et al*., 1996).

Equine infectious anaemia virus (EIAV), a member of the lentivirus subfamily of retroviruses (Stephens *et al*., 1986), affects members of the family *Equidae*, causing recurring cycles of fever and anaemia (Cheevers & McGuire, 1985). The only lentivirus known to be spread by an insect vector, the blood sucking horse fly (Foil *et al*., 1983), EIAV is readily transmittable between host animals and has a world-wide distribution. EIAV has the smallest and simplest genome of characterised lentiviruses, but nevertheless carries a gene, mapped within the *pol* region (Kawakami *et al*., 1987), encoding a 14.7 kDa dUTPase, consisting of 133 amino acid residues (Oroszlan *et al*., 1991; Roberts *et al*., 1991). The enzyme functions as a trimer (Bergman *et al*., 1995). dUTPase activity is essential for efficient viral infection of the natural host cells, the non-dividing equine monocyte-derived macrophages (Threadgill *et al*., 1993; Lichtenstein *et al*., 1995). dUTPase-deficient mutants of EIAV exhibit impaired replication in this cell type, where endogenous dUTPase activity is low (Steagall *et al*., 1995). However, in dividing fibroblastic cells, where there is a higher level of endogenous dUTPase activity, the virally encoded enzyme is not required for efficient propagation, as also found for the closely related feline immunodeficiency virus (FIV; Wagaman *et al*., 1993). Given the economic impact of infection of domestic animals, EIAV dUTPase is a potential target for antiviral drug design.

Amino acid sequence comparisons reveal that dUTPases contain highly conserved motifs (McGeoch, 1990). In the crystal structures solved so far of the *E. coli* (Cedergren-Zeppezauer *et al*., 1992), human (Mol *et al*., 1996) and FIV (Prasad *et al*., 1996) enzymes, these motifs appear to have similar roles. Structures of the *E. coli* (Larsson *et al*., 1996b) and human (Mol *et al*., 1996) enzymes in complex with substrate analogue and substrate, respectively, allowed the identification of the active site in which the conserved motifs are positioned. In the crystal structures of FIV and *E. coli* dUTPase, several residues at the C terminus of each subunit are not visible, whereas in the human enzyme complexes, almost all residues are well defined. None of these structures revealed the position of the divalent cation required for catalysis *in vivo*, presumably to be Mg^{2+} (Hoffman, 1987).

Here, we report the determination of the three-dimensional (3D) structure of EIAV dUTPase in unliganded and complexed forms, using the substrate analogue dUDP to characterise the active site. For the first time, a metal ion binding site has been identified in the active site of a trimeric dUTPase. The structure will serve as a starting point for site-directed mutagenesis experiments, which, together with the kinetic characterisation recently published (Nord *et al*., 1997), will provide mechanistic information.

Results and Discussion

Structure development with uARP

The structure of the EIAV dUTPase complex was determined by molecular replacement using the *E. coli* dUTPase crystallographic trimer as search model, the unliganded structure having proved resistant to this approach. The substantial differences in 3D structure between the enzymes from *E. coli* and EIAV (discussed below) gave rise to problems in the refinement, comparable to the difficulties reported in the structure determination of the FIV enzyme, where use of the *E. coli* model proved unsuccessful, necessitating employment of isomorphous replacement to solve the phase problem (Prasad *et al*., 1996).

Here, the program AMoRe (Navaza & Saludjian, 1997) gave a relatively clear molecular replacement solution using the *E. coli* model. Unfortunately, attempts to refine this starting model proved challenging. A straightforward application of stereochemically restrained refinement using the maximum likelihood based REFMAC program (Murshudov *et al*., 1997) combined with solvent updating through ARP (Lamzin & Wilson, 1997) were unsuccessful and did not lead to readily interpretable maps. In contrast, the use of the unrestrained option of REFMAC and ARP

(uARP, unrestrained automatic refinement procedure), with all starting atoms assigned as water molecules and all stereochemical restraints removed, worked well. The resulting maps (Figure 1) produced with minimal user intervention, had clear density for almost all of the structure. It was extremely straightforward to build a good EIAV dUTPase model, which could easily be refined with restrained protein stereochemistry imposed.

Quality of the models

Two structures are reported here, that of EIAV dUTPase at 1.9 Å, and that of the enzyme in complex with dUDP and Sr^{2+} at 2.05 Å resolution. The asymmetric unit contains one subunit with the three subunits of the trimer related by a crystallographic 3-fold axis in both space groups, *R*32 and $P4_132$.

Table 1 summarises the data collection and processing statistics, and Table 2 summarises the par-

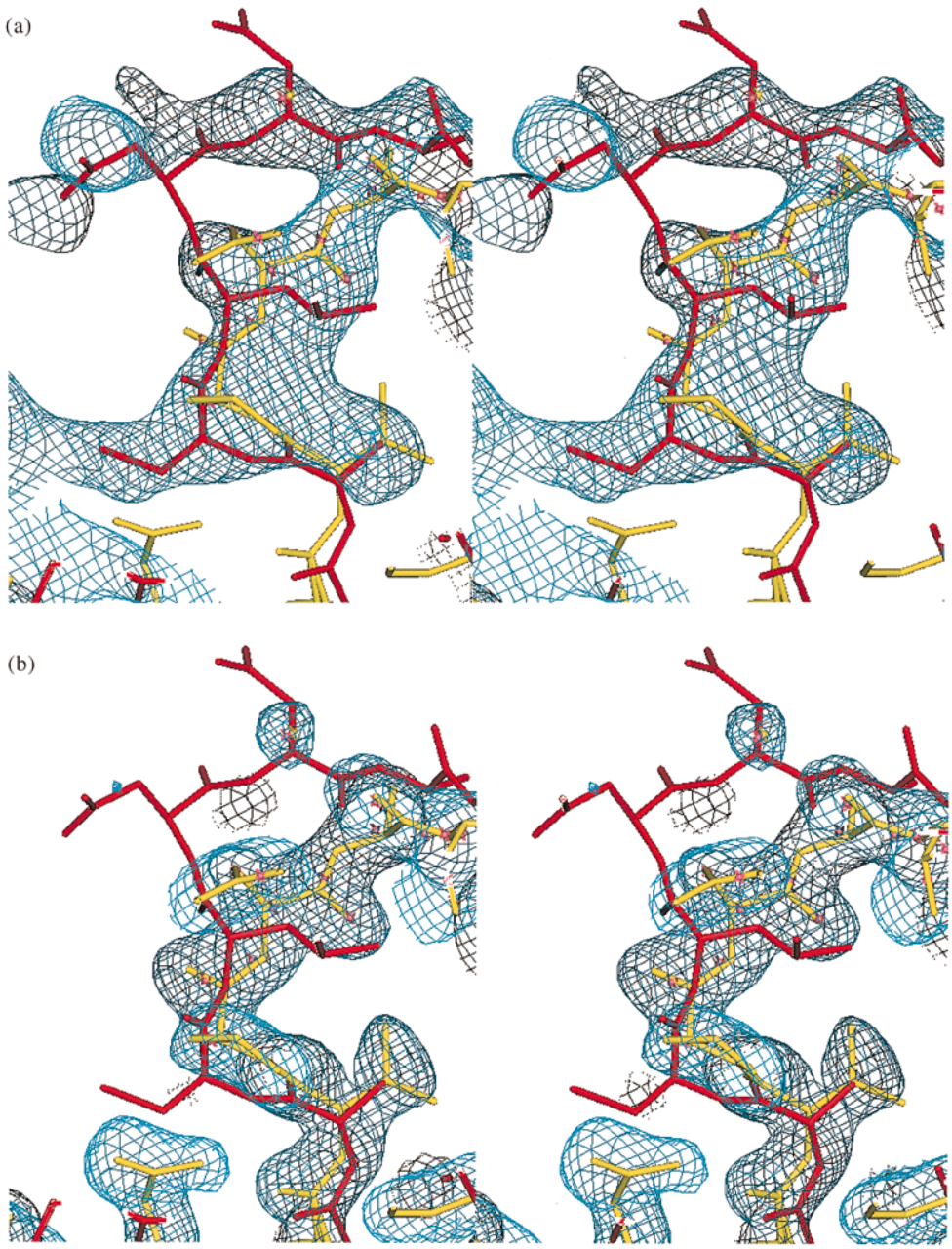

Figure 1. The improvement in the density maps produced by uARP for a loop on the surface of the enzyme. The coefficients are $2mF_o - dF_c$ as computed by REFMAC where *m* and *d* are maximum likelihood σA weights. (a) Initial map based on *E. coli* dUTPase as the molecular replacement model. (b) The same region after 60 cycles of uARP. The red models correspond to the initial and the yellow to the final refined model. The green atoms are the ''dummy water'' positions after uARP. The final density (not shown) is essentially indistinguishable from (b). Figures generated using QUANTA.

Table 1. Summary of data collection and processing for unliganded EIAV dUTPase and the complex with dUDP and Sr^{2+}

	Unliganded	Complex
Space group	*R*32	$P4_132$
Cell (Å)	$a = b = 86.27$ $c = 95.24$	106.7
Beam line, EMBL, Hamburg	X11	BW7B
Wavelength (Å)	0.916	0.885
Resolution (Å)	25 - 1.90	25 - 2.05
Total reflections measured	55,093	142,838
Unique reflections	10,717	13,598
Overall:		
Completeness (%)	97.5	100.0
$R(I)$ merge (%)	13.3	12.4
$I/\sigma(I)$	10.6	17.7
Highest shell (Å):	1.93 - 1.90	2.09 - 2.05
Completeness (%)	96.6	100.0
$R(I)$ merge (%)	42.2	68.6
$I/\sigma(I)$	3.6	3.8
B overall[a] ($Å^2$)	14.5	22.9

The merging $R(I)$ factor is defined as $\Sigma|I - \langle I\rangle|/\Sigma I$.
[a] Estimated from the Wilson plot (Wilson, 1942).

ameters of the final models. In the Ramachandran plot (Ramakrishnan & Ramachandran, 1965), there are no non-glycine residues in disfavoured regions in either of the structures. The estimated mean coordinate error according to the Data Precision Index (Cruickshank, 1996) as implemented in REFMAC, is 0.12 Å for the unliganded enzyme and 0.11 Å for the complex.

The recombinant enzyme subunit comprises 134 residues (Bergman *et al.*, 1995). The final electron density maps calculated with $2mF_o - dF_c$ coefficients and contoured at the 1σ level have unambiguous density for all main-chain atoms included in the model. However, there is no electron density for the C-terminal residues 121 to 134 in the unliganded enzyme, and 118 to 134 in the complex. The density disappears at a sharp point in the unliganded structure. In contrast, the chain adopts double conformations around residues 109-111 in the complex, and thereafter slowly becomes less well defined, suggesting disorder rather than chemical loss. Such C-terminal flexibility was also observed in the unliganded *E. coli*, FIV and human dUTPases. Zero occupancy was assigned to some atoms of three and six side-chains of unliganded and complexed enzyme, respectively, as they have poorly defined electron density. These side-chains generally belong to solvent-exposed residues with no crystal lattice contacts.

The quality of the electron density is high in most parts of the structures. Figure 2(a) shows a region around residues Gly68 and Gly69, in which there are two alternative main-chain conformations (discussed below). This region has a single conformation in the complex as seen in Figure 2(b). The well defined electron density for dUDP, a Sr^{2+} and its associated water molecules is shown in Figure 2(c). In the free enzyme there are 133 water molecules. A considerable number of those are positioned in the active site and several are expelled upon dUDP binding. There are only 80 water molecules in the model for the complex.

Alignment and superposition of known dUTPase models

The amino acid sequences of an extended set of homologous dUTPases have been aligned previously (Prasad *et al.*, 1996). Figure 3 shows a

Table 2. Summary of model properties

		Unliganded		Complex	
	σ	rms	No.	rms	No.
Distances (Å):					
Bond lengths (1-2 neighbours)	0.020	0.022	949	0.019	932
Bond angles (1-3 neighbours)	0.040	0.042	1285	0.040	1264
Dihedral angles (1-4 neighbours)	0.050	0.043	315	0.041	304
Planes (Å):					
Peptides	0.040	0.034	119	0.035	117
Aromatic	0.020	0.012	10	0.011	10
Chiral volumes ($Å^3$)	0.150	0.256	146	0.216	139
van der Waals contacts (Å):					
Single	0.30	0.20	375	0.19	351
Multiple	0.30	0.25	283	0.26	339
Possible H-bonds	0.30	0.11	25	0.15	87
Torsion angles (deg.):					
Planar	7.0	5.8	121	6.1	123
Staggered	15.0	16.1	175	13.4	170
Orthogonal	20.0	24.3	10	26.7	10
B-factors (Å):					
Main-chain bond	3.0	3.3	534	2.9	516
Main-chain angle	5.0	4.0	651	3.8	624
Side-chain bond	6.0	6.0	415	6.1	416
Side-chain angle	8.0	8.2	634	7.8	640

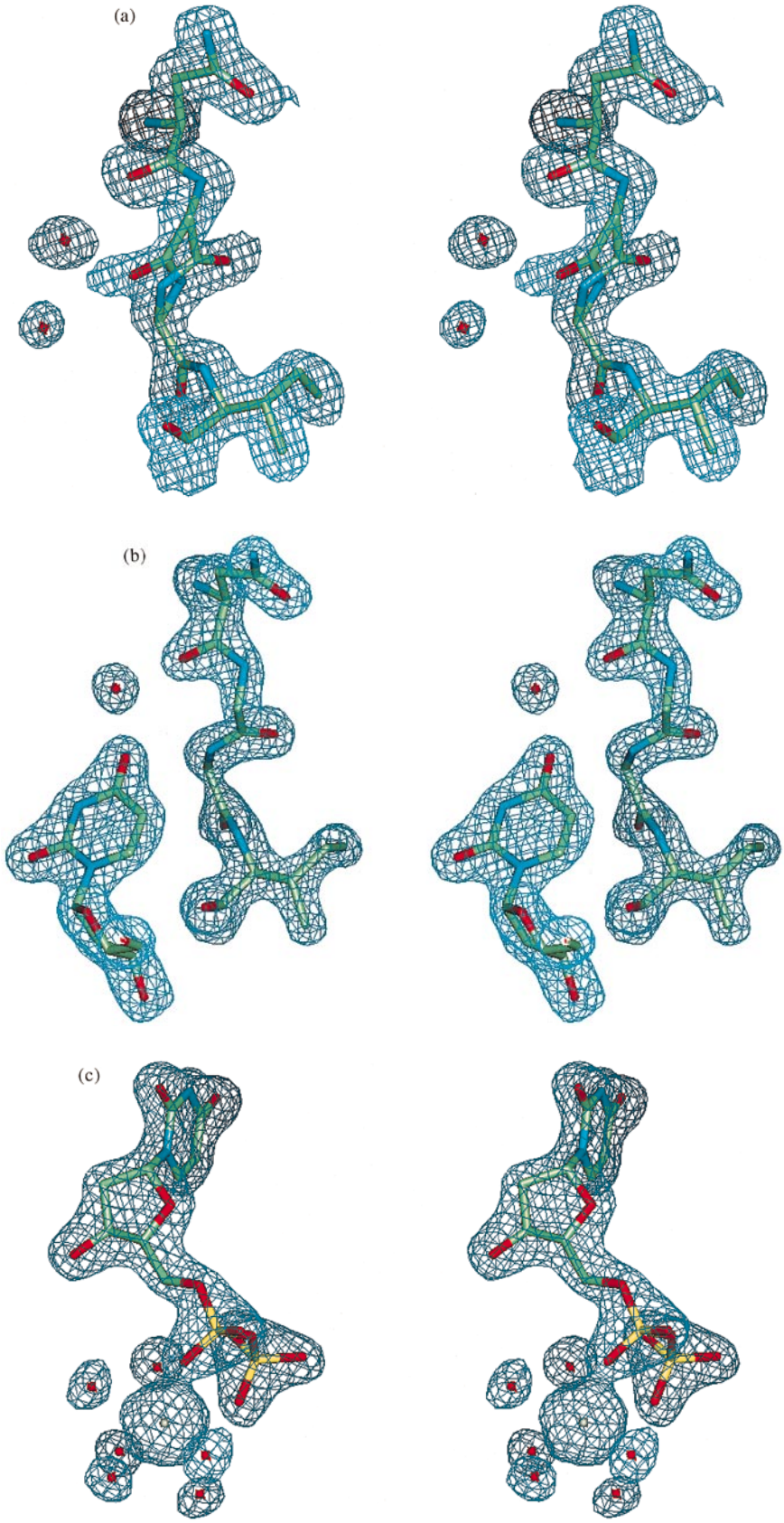

Figure 2. Stereo views of the electron density around the active site. (a) Unliganded EIAV dUTPase contoured at 1.2 σ showing the peptide Asn-Gly-Gly-Ile with the bond of Gly68-Gly69 in two alternative conformations (A and B) related by a peptide flip. (b) The Gly68-Gly69 peptide has a single, well-ordered orientation (B) in the complex contoured at 2 σ. (c) dUDP and Sr^{2+} (grey) with its six water ligands bound (red), again contoured at 2 σ. The bonds between the metal ion and the oxygen atoms of the α and β-phosphates are 2.47 Å and 2.69 Å, respectively. The bond lengths between the Sr^{2+} and its water ligands range between 2.77 Å and 2.85 Å. Figures generated using QUANTA.

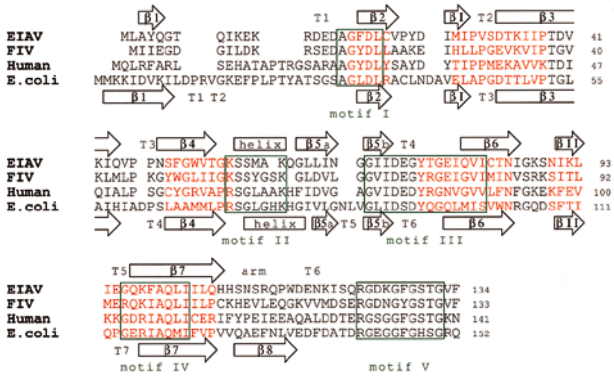

Figure 3. Amino acid sequence alignment of EIAV, FIV, *E. coli* and human dUTPases. The 53 structurally equivalent residues used in the superpositioning of the 3D structures are in red. The conserved sequence motifs defined previously (McGeoch, 1990) are boxed in green. The secondary structure elements and turns of the EIAV and *E. coli* enzymes are labelled.

revised alignment based on 53CA atoms identified from superpositioning of known 3D models of EIAV, *E. coli*, human and FIV dUTPase as structurally equivalent (see Materials and Methods for details). The 53 spatially equivalent CA atoms lie in five different regions along the sequence and their rms deviation is listed in Table 3. Despite limited sequence identity, the four enzymes exhibit strong conservation of the 3D structures in the core of the monomer. dUTPases contain five segments of conserved amino acid residues (denoted motifs 1-5; McGeoch, 1990), of which four are visible in the crystal structure of EIAV dUTPase, all located within or around the active site. The fifth motif is in the C-terminal region which is disordered in the crystals.

The architecture of EIAV dUTPase

The fold of the subunit, the quaternary organisation of the trimer and the location of the active site of EIAV dUTPase are all similar to those previously described for the *E. coli*, FIV and human enzymes. Therefore, these are not discussed in detail, but briefly summarised and differences between the other enzymes analysed.

The subunit

The main-chain H-bonding scheme of the subunit is shown in Figure 4, revealing the secondary structural elements. The core is predominantly composed of β-strands, forming two left-handed sheets (sheets I and II) in a distorted jelly roll. The only helical structure, a one-and-a-half-turn α-helix, lies on the surface of the enzyme. Nine loops, mainly regular β-turns, connect the strands. A β-hairpin containing turn 4 contributes to the formation of the active site. This hairpin has similar overall features in all known dUTPase structures, but the degree of sequence identity is low in this region, consistent with the involvement of the peptide backbone and water in substrate binding.

Figure 5(a) shows the backbone of the EIAV monomer onto which the other known dUTPase structures are superimposed (see Materials and Methods). The fold of the core of the subunit is highly conserved, with the two viral proteins being closest in overall fold. Differences are mainly situated in loops with a number of insertions and deletions. Extended loops on the surface of the *E. coli* and human enzymes, most clearly visible at the top of the Figure, reflect the insertions in these sequences. The extended arms show substantial differences, even for the viral pair, reflecting different quaternary structural interactions within the trimer; it should be remembered, however, that the superposition was only based on the residues in the core.

The trimer

The trimer has a triangular face of approximate dimensions 50 Å, and a height along the 3-fold axis, around which the three monomers associate, of 42 Å. Extending from the core, the C terminus of each subunit traverses the triangular face as a ''crossing'' arm, interacting with sheet I of the neighbouring subunit only through the formation of two main-chain H-bonds (Figure 4). Trp115 is buried in the adjacent subunit. In the *E. coli* enzyme, the interactions of the crossing arm are quite different in character, with numerous H-bonds to the neighbouring subunit as the arm participates as a β-strand in the jelly roll of this neighbouring subunit (Dauter *et al.*, 1998). Pro124, positioned where the crossing arm leaves the sub-

Table 3. rms deviations (Å) for the CA atoms of the available dUTPase structures aligned and superimposed as described in the text

Species	Monomer			Trimer		
EIAV (unliganded enzyme)	rms	Min.	Max.	rms	Min.	Max.
EIAV (dUDP-Sr^{2+}complex)[a]	0.21	0.19	0.46	0.39	0.34	0.83
FIV (unliganded enzyme)	0.67	0.62	1.28	0.95	0.89	1.58
E. coli (unliganded enzyme)	0.76	0.74	1.19	1.06	1.01	1.68
E. coli (dUDP complex)	0.78	0.74	1.30	1.06	1.00	1.74
Human (unliganded enzyme)	0.87	0.79	1.82	1.08	0.98	2.23

[a] The comparison to unliganded EIAV dUTPase is based on all atoms of the monomer.

unit core, has been attributed the function of a hinge (Bergdoll *et al.*, 1997) in the *E. coli* enzyme. The rigidity of the Pro, restricting the conformational freedom of the preceding residue, prevents the arm from folding back and interacting with its original subunit. Additionally, it orients the arm in a position favourable for interactions with the β-sheet of a neighbouring subunit with the arm running parallel to, and in the same plane as the sheet. The majority of dUTPases contain this hinge Pro, but in the EIAV enzyme the corresponding residue is a glutamine and the arm runs almost perpendicular to the adjacent β-sheet, nearly devoid of direct interactions. Pro114 positioned later in the sequence dictates the direction of the arm which ends up with the last visible residue pointing towards the active site. The deviations in the positions of the C-terminal arms of the four dUTPases with known 3D structures are evident in Figure 5(b), where the backbones of the four trimers are superimposed. The overall quaternary structure is again well conserved in the four enzymes.

The core of the trimer around the 3-fold axis

Several of the hydrophobic residues are buried in the jelly roll of the subunit. The others mainly participate in quaternary interactions around the 3-fold axis. There is a narrow channel along the 3-fold axis lined with side-chains predominantly contributed by sheet II of each subunit. The residues are positioned in alternating layers of hydrophobic and hydrophilic character, resulting in a ''pool'' of highly ordered water molecules enclosed at the centre (Figure 6(a)). The volume of the enclosed cavity, as calculated by SQUID (Oldfield, 1992), is 207 $Å^3$. In contrast, in the *E. coli* enzyme the interactions along the 3-fold axis are formed by closely packed hydrophobic residues, efficiently

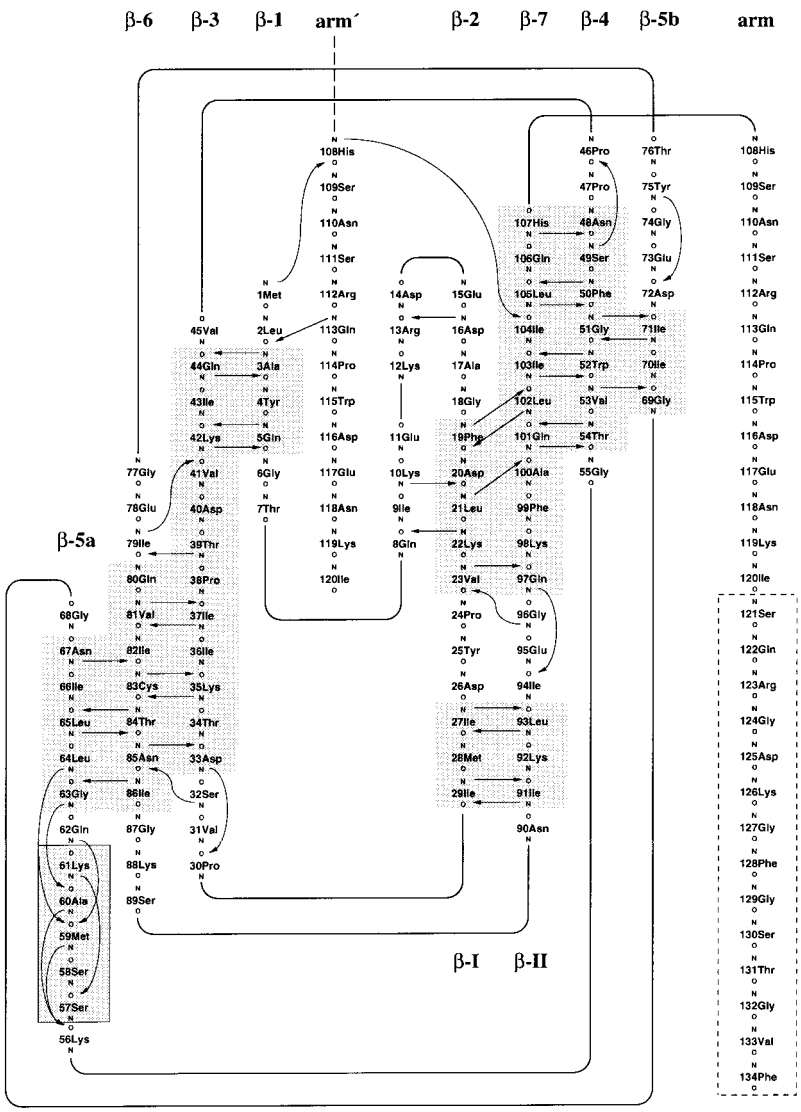

Figure 4. The main chain H-bonding scheme of EIAV dUTPase. The secondary structure elements are shaded and labelled. The flexible C terminus (residues 121-134) is boxed with a broken line. ''Arm'' denotes the ''crossing arm'' from a neighbouring subunit.

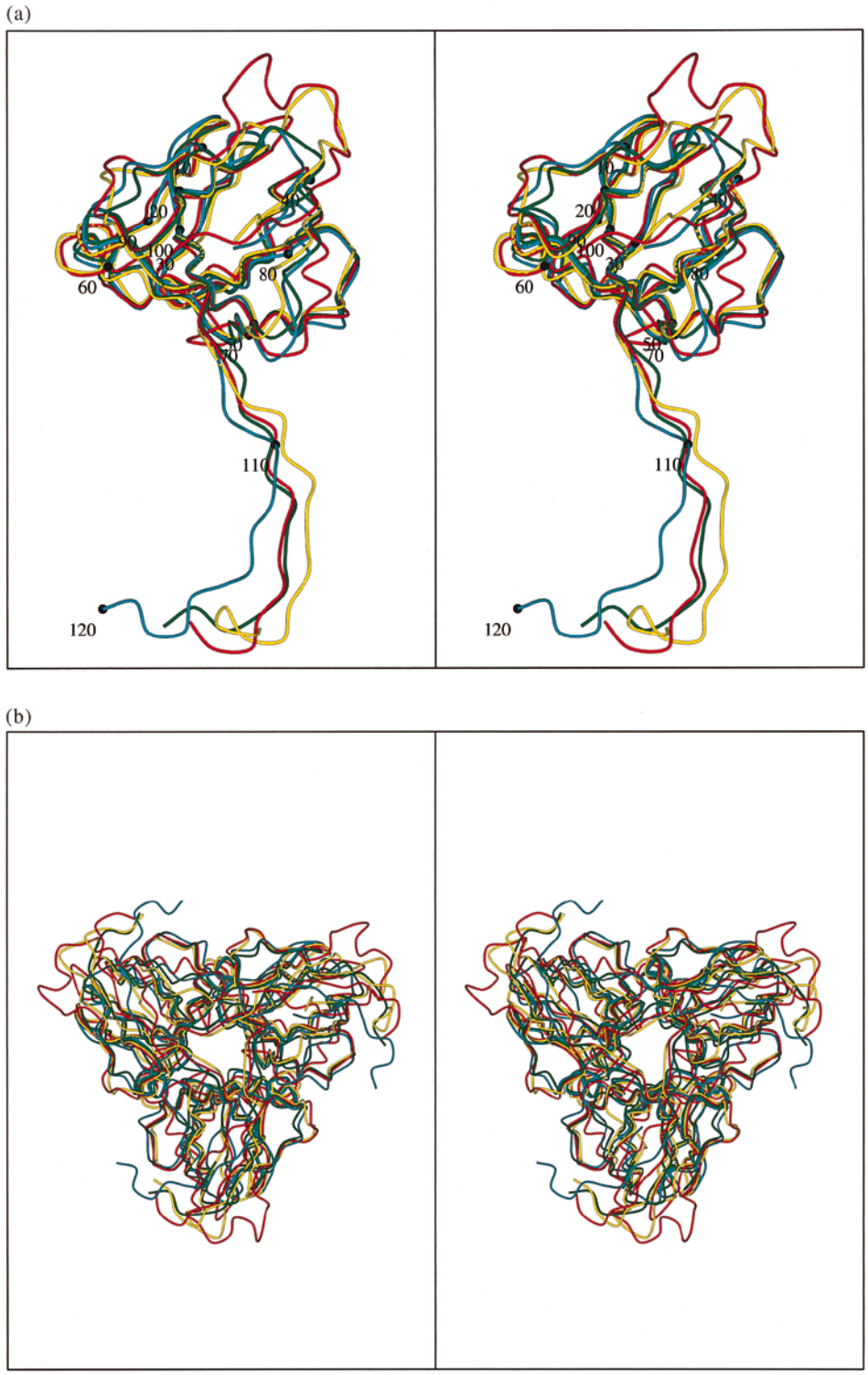

Figure 5. Superposition (based on the alignment of 53 CA atoms described in Figure 3) of the backbones of unliganded dUTPases from EIAV (blue), FIV (green), *E. coli* (red) and human (yellow), viewed down the 3-fold axis. (a) Monomers. Every tenth residue of EIAV dUTPase is numbered. (b) Trimers. 159 CA atoms were superimposed. Figures were generated using MOLSCRIPT (Kraulis, 1991).

excluding water (Figure 6(b)). The interactions along the 3-fold axis in the FIV enzyme is of an intermediate character, with a smaller pool of water molecules located at the central part of the core (Figure 6(c)), while in the human enzyme water molecules present around the 3-fold axis are not completely buried as they have contact with bulk solvent *via* the open triangular face of the trimer (data not shown). Thus, while the overall subunit arrangement is highly conserved between the four enzymes, the interactions providing its stability vary substantially.

In both the EIAV and the *E. coli* enzymes, the side-chains of three symmetry-related tryptophan residues point towards the 3-fold axis. In the viral enzyme, the rings are positioned perpendicular to one other, interacting through edge-to-face contacts, one of the preferred orientations for aromatic-aromatic interactions (Burley & Petsko, 1985; Hunter & Sanders, 1990), and are surrounded by several water molecules. The Trp rings in the *E. coli* enzyme are positioned with a face-to-face geometry with the aliphatic portion of three arginine side-chains sandwiched between them, creating a hydrophobic barrier for the exclusion of water.

Active site

The dUDP-Sr^{2+} complex

The folds of the protein in the unliganded and complexed forms of the EIAV enzyme are very similar, with only minor conformational adjustments occurring to accommodate the nucleotide and metal ion. Superposition of all CA atoms gives an rms deviation of 0.21 Å (the error in co-ordinates is 0.12 Å). Even though the enzyme was cocrystallised with dUDP and Sr^{2+}, the C-terminal residues remain disordered in the complex. Enzyme from redissolved crystals retains full enzy-

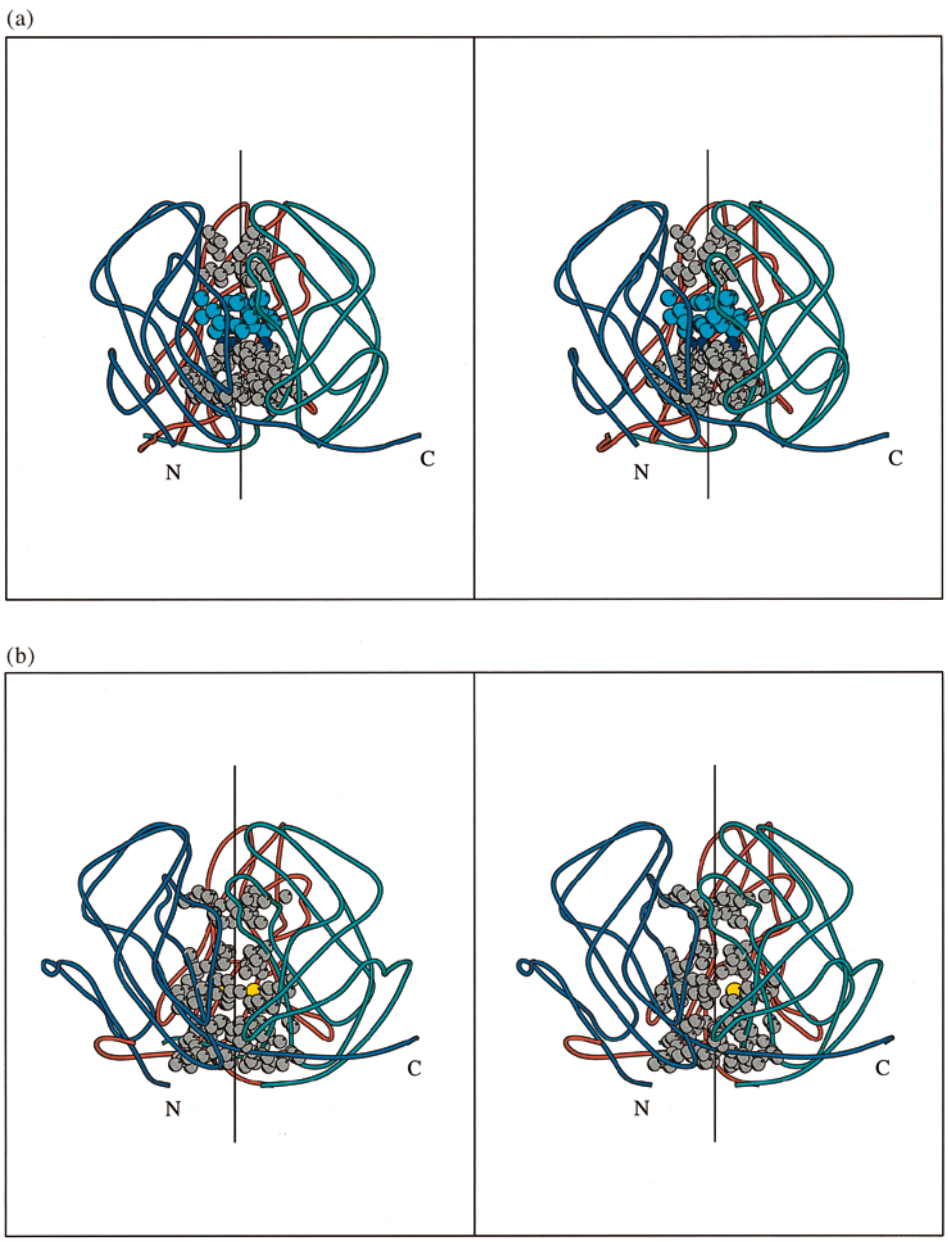

Figure 6 (a)-(b) *(legend overleaf)*

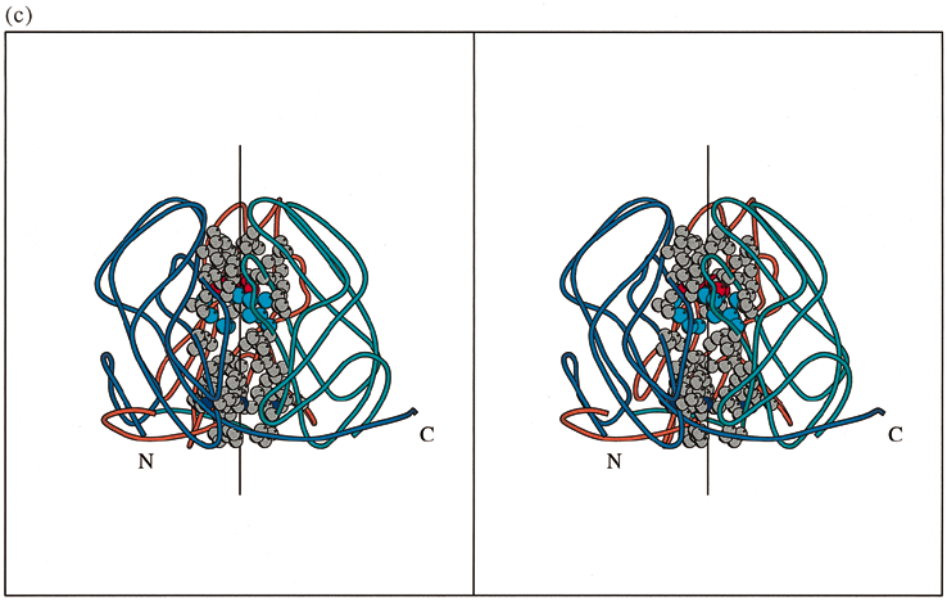

Figure 6. dUTPase trimers viewed perpendicular to the 3-fold axis. (a) The EIAV dUTPase has an extensive pool of ordered water molecules in the central cavity, isolated from bulk solvent by closely packed groups of hydrophobic side-chains above and below. (b) The centre of the 3-fold interface in *E. coli* dUTPase is totally hydophobic, with no buried water molecules. (c) The FIV enzyme has a small pool of water molecules. Colour code: non-polar residues, grey; oxygen, red; nitrogen, blue; sulphur, yellow; water, light blue.

matic activity, and as removal of C-terminal residues is associated with loss of activity (Shao *et al.*, 1997; Vertessy, 1997), the full-length protein is assumed to be present in the crystals. In the human enzyme, the flexible C terminus interacts with the substrate (Mol *et al.*, 1996).

The dUDP and Sr^{2+} bind in the active site, whose location was established in earlier studies. The positions of all atoms of the nucleotide and the ligands of the Sr^{2+} are well defined (Figure 2(c)). dUDP binding occurs in shallow depressions at the subunit interfaces. Conserved residues from all three subunits have been described to partake in the formation of each complete active site (Mol *et al.*, 1996). One subunit provides the residues necessary for base and sugar recognition, the β-hairpin. A second is involved in phosphate binding, while the third, not seen in the present structure, is presumed to cover the active site during catalysis, as in the human enzyme.

Binding of the uracil

Uracil recognition and sugar specificities are mediated by a β-hairpin (residues 67-82) formed by strands β-5b and β-6, conserved in other dUTPases. A tyrosine at the scaffold of the β-hairpin and a turn of type I are characteristic features of the uracil-binding site.

Binding of the uracil ring is achieved through four H-bonds between backbone atoms, the ring and one water molecule (Figure 7(a) and (b)). The main-chain amide and carbonyl groups of Gln80 are bound to O2 and N3 of the uracil ring, respectively. The O4 of the ring forms H-bonds to the amide of Gly69 and to a structurally conserved water molecule. As this buried water is also H-bonded to the main-chain amide of Ile82 and to the main-chain carbonyl of Asn67, it serves the dual roles of knitting the strands of the β-hairpin together as well as participating in substrate binding. This conserved water is present in the native structure as well, and has been observed in all dUTPase structures determined so far.

Conformational change

Binding of the uracil ring induces a local conformation change, a flipping of the peptide backbone affecting residues Gly68 and Gly69 (Figure 2(a) and (b)). Such a flip has not been observed before in a dUTPase. The structure of the closely related FIV dUTPase in the unliganded form assumes the conformation here determined for the EIAV enzyme-dUDP-Sr^{2+} complex. In the unliganded EIAV dUTPase, the carbonyl of Gly68 and the amide group of Gly69 have two conformations, A and B, both partially occupied. In the B conformation, O68 points away from the β-hairpin, accepting hydrogen atoms from two water molecules located close to the pool of water enclosed around the 3-fold axis. Meanwhile, N69 is directed into the water-containing β-hairpin, donating its proton to a water with partial occupancy (Figure 2(a)). In the A conformation, neither O68 nor N69 have any H-bonding partners. Upon complex formation, the backbone comprising the two glycine residues flips over if previously in confor-

mation A. As the O4 of uracil acts as an H-bond acceptor, the peptide flip enables the formation of an additional H-bond to the dUDP as it positions the amide of Gly69 within H-bonding distance of the O4. Formation of this bond will not be possible if the two glycine residues maintain an A conformation, where the carbonyl group of Gly68 will be directed towards the O4, unable to form an H-bond.

The catalytic specificity of EIAV dUTPase towards the base is not as stringent as that of the *E. coli* enzyme (Nord *et al*., 1997). The presence of a bulky methyl group in the C5 position of the thymine ring ought to cause a steric clash with the β-hairpin wall upon binding to EIAV dUTPase. However, a certain amount of dTTP binding and hydrolysis occurs, probably enabled by the mobility and the small size of the two glycine residues (Gly68-Gly69). In human dUTPase, an alanine is inserted between the corresponding glycine residues (Figure 3), but reports on the amount of dTTP hydrolysis exhibited by this enzyme are contradictory (Williams & Cheng, 1979; Mol *et al*., 1996). The *E. coli* dUTPase has, however, been demonstrated to discriminate efficiently against dTTP (Larsson *et al*., 1996a). This is probably due to the insertion of three additional residues (compared with the EIAV enzyme) in the part of the β-hairpin interacting with the O4 and C5 of uracil, creating a highly strained backbone conformation (Figure 7(c)). As the side-chain of Asn84 of the bacterial enzyme is directed towards the uracil ring,

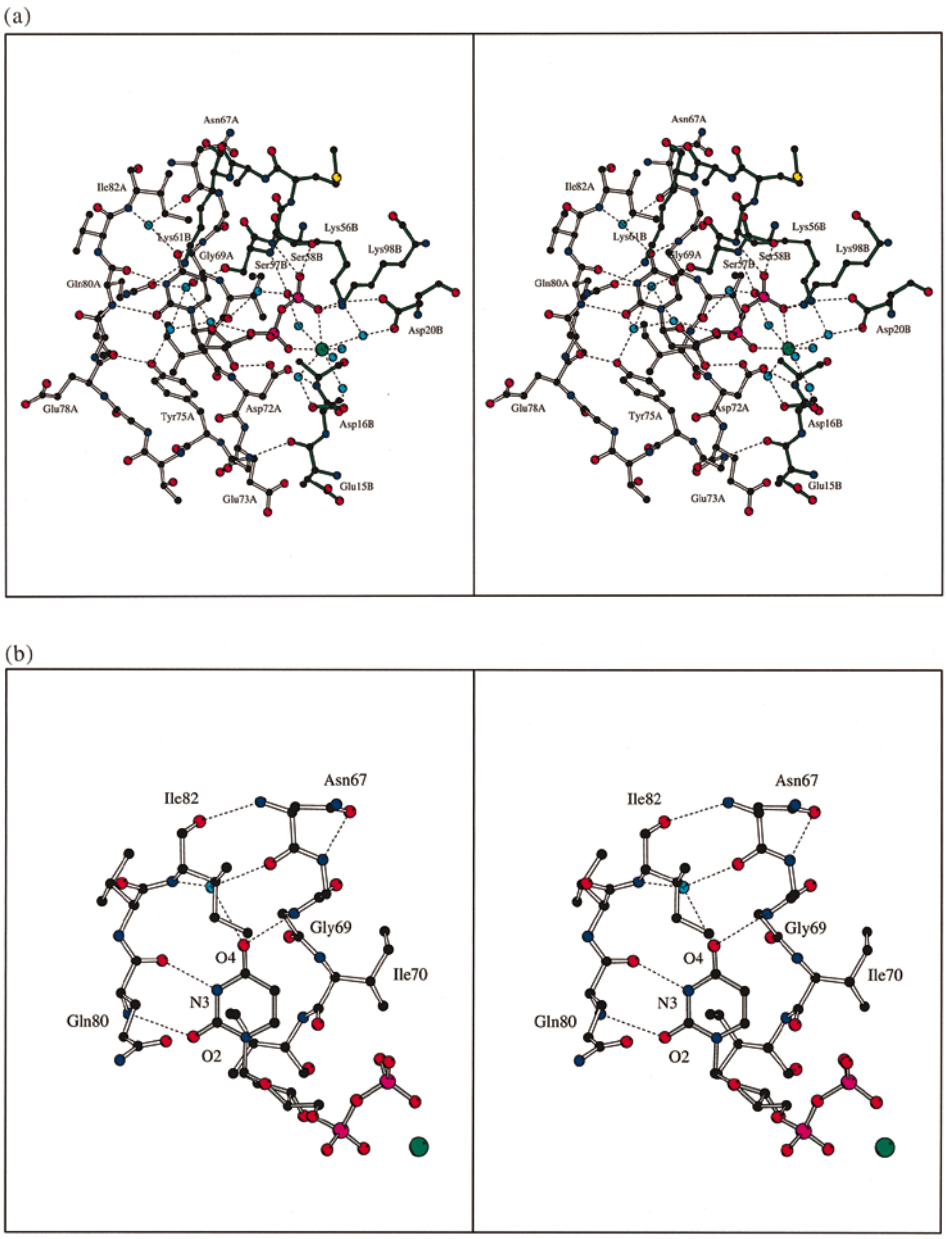

Figure 7 (a)-(b) *(legend overleaf)*

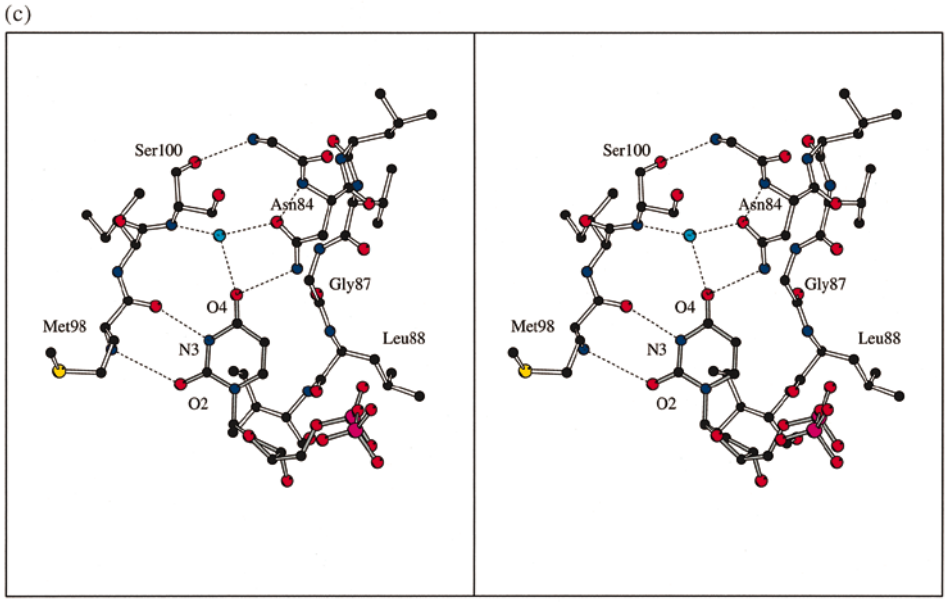

Figure 7. (a) An overall view of the active site of EIAV dUTPase illustrating the structural basis of substrate specificity and the H-bonding pattern between substrate analogue and enzyme. The bonds of one subunit are in white and those of the symmetry related subunit in a trimer are in green. The inhibitor bonds are grey. Water molecules are shown in blue. (b) Close-up of uracil binding in EIAV dUTPases. The structural water which binds the uracil O4 atom to the main-chain of residues Asn67 and Ile82 is seen at the top. (c) Uracil binding in *E. coli* dUTPase. Figures were generated using MOLSCRIPT (Kraulis, 1991).

forming an H-bond with O4, limited space is available and dTTP binding is prevented. Inspection of the model of EIAV dUTPase suggests that binding of cytosine would result in an impaired H-bond pattern compared with that of uracil binding. Additionally, the position of the structurally conserved water might be perturbed when the O4 of uracil is replaced with the amino group of the cytosine, possibly affecting the structural role of this water. Of the other naturally occurring nucleotide bases, the purine bases are easily excluded due to their large size.

Deoxyribose interactions

Apart from mediating uracil specificity, the β-hairpin, through the creation of a pocket, renders dUTPases specific with regard to the pentose moiety of the substrate. The aromatic ring of the conserved Tyr75, positioned in this pocket, is stacked with the deoxyribose ring, efficiently preventing ribose binding. Chemical modification of this tyrosine results in a complete loss of enzyme activity (Vertessy *et al.*, 1996), consistent with its importance in substrate binding. The OH of Tyr75 is anchored to the wall of the β-hairpin through an H-bond, and its position is further stabilised by a turn (residues 72-75). On the opposite side of, and in van der Waals contact with, the deoxypentose is Ile71. Accommodation of a ribose is, therefore, implausible as the pentose would interfere with Tyr75. Having a tyrosine positioned in the deoxyribose recognition pocket seems to be favourable, as this tyrosine is almost totally conserved among dUTPases. One exception is mouse mammary tumour virus (MMTV) dUTPase, in which tyrosine is replaced by phenylalanine, still enabling the stacking of deoxyribose but making the anchoring of the ring to the hairpin wall impossible. This replacement, with concomitant enhanced mobility of the aromatic ring, could allow increased binding of UTP. A decreased discrimination between dUTP and UTP has indeed been reported for MMTV dUTPase (Björnberg & Nyman, 1996).

The 3′-hydroxyl group of the deoxyribose accepts an H-bond from the main-chain amide of the strictly conserved residue Asp72, and donates an H-bond to the side-chain carboxyl group of the same aspartate. The interactions between the side-chain of Asp72 and the 3′OH of the deoxyribose are identical with those reported for the *E. coli* and human enzymes (Larsson *et al.*, 1996b; Mol *et al.*, 1996). For the bacterial enzyme, it has been suggested that a water molecule H-bonded to the corresponding aspartate might be involved in nucleophilic attack on the α-phosphate occurring during catalysis. Through the formation of an H-bond to the main-chain amide of Ala17 of a neighbouring subunit, the side-chain of Asp72 in EIAV dUTPase also participates in subunit association, an interaction not seen in the bacterial enzyme.

Phosphate interactions

The phosphate portion of dUDP interacts almost exclusively with residues provided by a subunit adjacent to the β-hairpin forming subunit (Figure 7(a)), the interactions being either direct or

mediated by well-defined water molecules. The very short α-helix is positioned close to the phosphate moiety of the dUDP. The α-phosphate, where nucleophilic attack occurs, is H-bonded through water bridges to the side-chains of Asp26 and Glu95 and surrounded by several water molecules. The α-phosphate has no direct H-bonds to protein atoms, in contrast with the *E. coli* enzyme (Larsson *et al*., 1996b). The β-phosphate accepts H-bonds from the main-chain amide groups of residues 57 and 58, and from the side-chain hydroxyl of Ser58. Additionally, the positively charged side-chain amino groups of Lys56 and Lys98 interact with the β-phosphate. The H-bond pattern for the β-phosphate moiety is in agreement with that seen in the model of the human enzyme-dUDP complex (Mol *et al*., 1996). In the human enzyme-dUTP complex, binding of the γ-phosphate results in the rotation of an arginine side-chain interacting with the β-phosphate in the dUDP complex. The position vacated by the rotation is occupied by the γ-phosphate, allowing it to interact with another arginine side-chain. The corresponding residues in EIAV dUTPase are two lysine residues, and a scenario where the side-chain of Lys56 rotates to allow the entrance of a γ-phosphate is plausible with regard to available space. Lys98 is positioned so that it could easily interact with and neutralise the charge on an incoming γ-phosphate.

dUDP is a moderately strong inhibitor of EIAV dUTPase (Shao *et al*., 1997), but is not hydrolysed to any notable extent by the enzyme (Nord *et al*., 1997). The absence of the γ-phosphate, present in the true substrate, probably prevents the flexible C terminus from becoming correctly positioned, thereby impeding the formation of the necessary transition state. The inability of the C-terminal arm to position itself in the absence of a γ-phosphate has been confirmed in recent spectroscopic studies (Vertessy *et al*., 1998). As the differences between the analogue and the substrate are minor, the dUDP model, nevertheless, provides a reasonable starting point for discussing enzyme-substrate interactions.

The metal site

Catalysis requires Mg^{2+} (Nord *et al*., 1997; Shao *et al*., 1997) and the Michaelis complex is assumed to involve Mg^{2+} and dUTP. Kinetic studies using dUTPαS (2′-deoxyuridine 5′-(α-thio)triphosphate) has provided evidence for the involvement of the α-phosphate in metal ion binding during catalysis (Bergman, 1997). It has been proposed that one or two magnesium ions bind to the substrate during catalysis; one between the α and the β-phosphates, and possibly another one at a second site. However, no metal ion sites were found in the active site in previously determined dUTPase structures. In the crystal structure of FIV dUTPase, a Mg^{2+} was observed, not in the active site, but on the 3-fold axis (Prasad *et al*., 1996).

Here, Sr^{2+} was chosen instead of Mg^{2+} as it is easier to detect in electron density maps and, in addition, it readily cocrystallises with the enzyme and dUDP. The position of the Sr^{2+} in the dUDP-liganded enzyme indicates experimentally, for the first time, a likely approximate location of a metal ion in the active site. Sr^{2+} co-ordinates to the diphosphate moiety of dUDP, i.e. to one oxygen on each of the two phosphates, and to six well ordered water molecules. Bond lengths between the ion and its co-ordinating water molecules and phosphate oxygen atoms are similar to those observed in crystal structures of strontium hydroxide octahydrate and strontium diphosphate, respectively (Smith, 1953; Hagman *et al*., 1968). The Sr^{2+} co-ordination sphere makes contact with the side-chains of Arg13, Asp20 (a conserved residue), Lys56 and Glu95, all belonging to the subunit involved in phosphate binding (Figure 7(a)). There is a second shell of water ligands at 4-6 Å.

Due to the differences in size between Sr^{2+}, with an ionic radius of 1.13 Å, and Mg^{2+}, with 0.65 Å, the presence of a Sr^{2+} clearly affects the detailed geometry of the active site. Thirty per cent of the Mg^{2+} activity is retained with Sr^{2+} (Figure 8), indicating that the Sr^{2+} detected in the structure closely mimics the position and role of Mg^{2+} in the fully active enzyme. The general structure of the active site, the binding of the substrate analogue and the H-bond pattern of the β-phosphate, in particular, appear to be in good agreement with the human dUTPase-dUDP complex which has no metal bound (Mol *et al*., 1996), suggesting that the size of the ion does not substantially distort the local environment. Ordering of the C-terminal residues, suggested to cover the substrate during catalysis, may be perturbed by the size of the Sr^{2+}. Alternatively, as dUDP is an inhibitor rather than a substrate, the C terminus might not be ordered even with magnesium present. As the C-terminal arm is

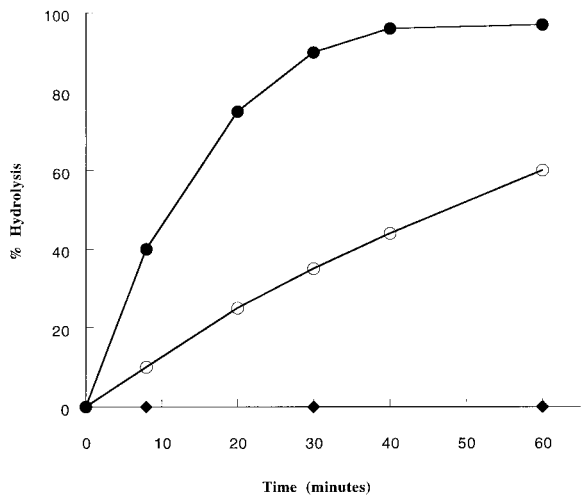

Figure 8. The activity of EIAV dUTPase in the presence of Mg^{2+} (●), Sr^{2+} (○) and no metal ion (◆) respectively. Assay conditions: 3 nM enzyme 0.3 mM dUTP and 10 mM metal ion.

disordered in the *E. coli* dUTPase-dUDP complex, which lacks a visible metal ion (Larsson *et al*., 1996b), and as spectroscopic studies show that the γ-phosphate is required for binding of the arm (Vertessy *et al*., 1998), the latter seems most probable.

In EIAV dUTPase, there is a preponderance of charged amino acid side-chains and water molecules in and around the active site. A cluster of positively charged residues is positioned where the γ-phosphate is proposed to bind, facilitating binding by complementing the negative charges on the phosphate group. A second charged cluster including several negatively charged residues (Glu15, Asp16, Asp72, Glu73), could participate in metal ion binding during catalysis (Figure 7(a)). The well-ordered water molecules in the active site of the dUDP complex can be explained by the presence of numerous charged amino acid side-chains and by the water co-ordinating abilities of Sr^{2+}.

Conclusions

The 3D structure of EIAV dUTPase in complex with the substrate analogue and inhibitor dUDP and a metal ion, Sr^{2+}, was solved using the *E. coli* enzyme as a molecular replacement model. The refinement was not straightforward, and required a combination of maximum likelihood minimisation in the program REFMAC and appropriate weights for the resulting maps with an unrestrained automated replacement procedure, uARP. This resulted in greatly improved phases and electron density maps, from which the structure could easily be rebuilt. The structure of the complexed enzyme was subsequently used for solving of the structure of the unliganded enzyme. The approach for structure determination should be of general applicability, and has been used in the determinations of the structures of glucoamylase and laccase (Sevcik *et al*., 1998; Ducros *et al*., 1998).

Features of structural similarities between the EIAV dUTPase and other dUTPase structures include the homotrimeric arrangement of subunits, the presence of an intersubunit traversing arm, a flexible C terminus, and the location and overall design of the active site. Differences related to trimer assembly are reflected by the nature of interactions along the 3-fold interface. The EIAV, FIV and human enzymes all contain bound water inside the central channel. In contrast, the *E. coli* enzyme displays a set of closely packed hydrophobic side-chains efficiently excluding water, rendering it an exception amongst known structures. The number of H-bonds between the crossing arm and the adjacent subunit differs between the EIAV and the bacterial dUTPases. While the crossing arm in the bacterial enzyme is a β-strand forming numerous interactions where it is embedded in the adjacent monomer, the arm in the viral enzyme is more hydrophilic and traverses the triangular face of the trimer as a coiled structure, differently directed across the surface. The overall features of the monomer is conserved with regard to other dUTPases as can be expected from sequence similarities. Differences in the lengths of the polypeptides of the enzymes are generally reflected as changes in the structure of surface loops, affecting particularly the N-terminal region. Among the residues selected for superimposing the structures only a few were conserved, but these conserved residues are inherently involved in substrate recognition or have critical structural roles.

Detailed examination of the active sites, which are positioned in clefts at subunit interfaces in accordance with previously described dUTPase structures, reveals significant differences between the EIAV and the *E. coli* enzymes. This is especially pronounced in the disparate distribution of charged residues, but also concerns the mode of interactions with dUDP. A peptide backbone flip involving two glycine residues is utilised in the viral enzyme for uracil binding, and presumably renders the EIAV enzyme more flexible in this region than the bacterial enzyme which has a highly strained conformation in this area due to the presence of additional residues. The pattern of interactions with the O4 of the uracil is of dissimilar character in the two enzymes, as the viral enzyme employs backbone interactions for binding whereas the bacterial enzyme uses side-chain interaction. Binding of the deoxypentose moiety is similar in the two enzymes, while binding of the phosphate moiety differs. The *E. coli* dUTPase mainly interacts with the α-phosphate of dUDP while the EIAV enzyme has its major interactions with the β-phosphate, this difference is possibly caused by the presence of a metal ion in the latter structure. Binding of the phosphate moiety in the EIAV enzyme seems to be in good agreement with the binding of the corresponding moiety of dUDP in the human enzyme.

The contribution of individual residues to catalysis still remains to be elucidated and requires the crystallisation of other ligand/metal complexes in which the arm resides in an ordered conformation. Catalysis requires a divalent metal ion *in vivo* presumably Mg^{2+}. Substitution of Mg^{2+} by Sr^{2+} allows the enzyme to maintain partial activity, and thus the Sr^{2+} in the present structure is a good indicator for the position in the active site, of a catalytically essential metal ion not seen in previously described dUTPase structures.

Materials and Methods

Enzyme preparation

Recombinant EIAV dUTPase, containing an additional Met at the N terminus, was purified from an overproducing *E. coli* strain as described by Bergman *et al*. (1995). The pooled chromatographic peak fractions containing approximately 30 mg of pure active enzyme were dialysed extensively against 10 mM 3-[*N*-morpholino]-propanesulphonic acid (Mops, pH 7.2), containing 1 mM

β-mercaptoethanol, and then concentrated to 20-45 mg/ml prior to crystallisation.

Activity measurements with Sr^{2+}

To determine the level of enzyme activity retained when Mg^{2+} is substituted by Sr^{2+}, the conversion of dUTP to dUMP at 25 °C was monitored by FPLC as described by Björnberg & Nyman (1996). The enzyme concentration was 3 nM, and that of the metal ion 10 mM for either Sr^{2+} or Mg^{2+}. Altering the concentration of Sr^{2+} did not increase the level of hydrolysis (data not shown). The reaction buffer used was 50 mM Tris-HCl (pH 7.4) containing 1 mM DTT. Hydrolysis was initiated by the addition of dUTP to a final concentration of 0.3 mM, the final volume of the sample being 500 µl. Aliquots of 25 µl were withdrawn from the reaction mixture every 15 minutes and quenched by the addition of EDTA to a final concentration of 45 mM. Each sample was diluted with 1.5 ml of 50 mM Tris-HCl (pH 8.0), and injected onto a MonoQ column (HR 5/5 Pharmacia) attached to an FPLC system. The hydrolysis products were eluted with a linear gradient of NaCl in 50 mM Tris-HCl (pH 8.0). The hydrolysis level caused by background Mg^{2+} in the water and buffer substances was estimated using a reference sample where the metal ion solution was replaced by water. The level of hydrolysis contributed by Mg^{2+} present in the strontium salt was calculated from the manufacturer's (Merck, Darmstadt) analysis of Mg^{2+}. In addition, a reference sample with the corresponding Mg^{2+} concentration (0.1 µM) was assayed for activity, but no activity could be detected.

Crystallisation

Crystals were obtained as described (Persson *et al.*, 1997). Briefly, crystals were grown at 4 °C using sitting drop vapour diffusion with 10 µl drops and 1 ml reservoir solutions. Equal volumes of protein and reservoir solutions (0.1 M imidazole malate and PEG) were mixed. The conditions for the crystal used for data collection on the unliganded enzyme were 1.4 µg protein/µl after mixing, pH 7.0 and 15 % (v/v) PEG3400. It had the shape of a rhombohedron and required a growth period of five months to reach a size of 0.25 mm × 0.25 mm × 0.15 mm. The crystal of the complex was obtained by including 20 mM $SrCl_2$ and 5 mM dUDP in the drop, and the enzyme concentration was 3 µg/µl after mixing. The precipitant was 21 % (v/v) PEG400. Crystals appeared after a fortnight, had rhombic dodecahedral morphology and a maximum edge size of 0.25 mm.

X-ray data collection and processing

X-ray intensities for unliganded EIAV dUTPase and its complex with Sr^{2+} and dUDP were collected at 4 °C from single crystals at the EMBL outstation, DESY, Hamburg, using a MarResearch imaging plate scanner and processed with the HKL suite (Otwinowski & Minor, 1997; Table 1). The asymmetric unit in the native R32 crystal contains one monomer of 14,678 Da with a specific volume of 2.32 $Å^3$/Da, equivalent to 47 % solvent (Matthews, 1968). The asymmetric unit of the $P4_132$ crystal of the complex also contains one monomer with a specific volume of 3.43 $Å^3$/Da, giving a solvent content of 64 %. The merging R factor and the $I/\sigma(I)$ ratio are shown as a function of resolution in Figure 9.

Molecular replacement

All programs used were from the CCP4 suit (Collaborative Computational Project, Number 4, 1994). The structure of the complexed enzyme was solved by molecular replacement using the program AMoRe (Navaza & Saludijan, 1997) with the trimer of *E. coli* dUTPase (1DUP, Protein Data Bank; Bernstein *et al.*, 1977) as a search model. Attempts to solve the structure of the unliganded enzyme were unsuccessful. The difference in packing density, V_M, for the two crystal forms may be important here. The closer intermolecular packing of the R32 form may make the molecular replacement more difficult for the latter, with poorer discrimination between inter- and intramolecular vectors. Use of the refined native EIAV model with the EIAV native data gave a clear and obvious solution after the analysis was complete.

Analysis of the complex led to a clear correct solution. The rotation function gave a series of solutions with very similar correlation coefficients. Three of them were related by 120 ° rotation about the [111] direction of the cell, as expected for the crystallographic trimer. The translation function and rigid body fitting (in $P4_132$, but not $P4_332$) confirmed the correctness of these, giving a correlation of more than 34.2 % compared with 19 % or less for the incorrect ones. The R factor for the correct solutions was 52.7 %, for the incorrect ones higher than 55.5 %.

The structure of the unliganded enzyme was subsequently solved in a straightforward manner by molecular replacement using the refined model of the complex.

Unrestrained ARP

Initially, the model corresponding to one of the three symmetrically equivalent molecular replacement solutions was subjected to restrained crystallographic refinement using the program REFMAC (Murshudov *et al.*, 1997) combined with the ARP procedure for automatic selection of waters (Lamzin & Wilson, 1997). However, attempts with variants of this model from which selected sets of side-chains had been removed, gave essentially no reduction in R_{free} from its initial value of 55 % and did not lead to more easily interpretable electron density maps.

Therefore, an alternative approach was used: the unrestrained option of ARP has recently been successful in several applications (Sevcik *et al.*, 1998; Ducros *et al.*, 1998), producing excellent and easily interpretable electron density maps with consequent substantial savings in time devoted to visual interpretation of intermediate maps. All atoms in the initial model are treated as solvent and no stereochemical restraints applied. After every cycle, a selected number of atoms in the weakest electron density or those which moved closer than 0.7 Å to other atoms were discarded. The rejected atoms can include protein atoms which do not correspond to real features in the new structure. A number of new atoms are selected from the difference density peaks at an acceptable distance from existing atoms.

All the atoms of the *E. coli* dUTPase (1DUP) starting model were changed to water molecules. The resulting model was subjected to unrestrained refinement using

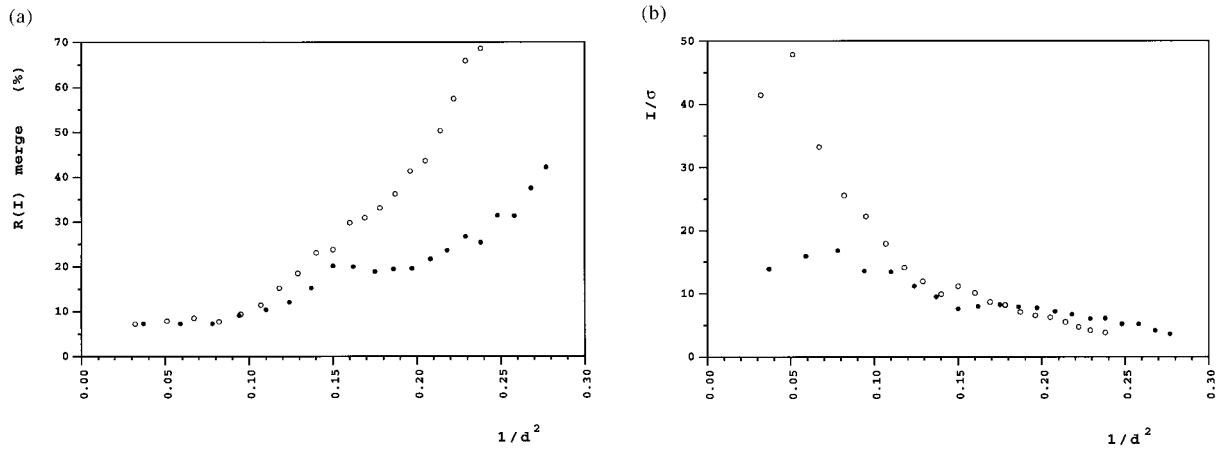

Figure 9. (a) The merging $R(I)$ factor and (b) $I/\sigma(I)$ as a function of resolution for unliganded dUTPase (filled circles) and its complex (open circles).

REFMAC combined with ARP which was allowed to reject or accept up to 20 atoms in each cycle. Four jobs were run, the first two with ten cycles and the last two with 20 cycles. The progress of R and R_{free} and the number of atoms in the model in successive cycles are shown in Figure 10. After the first ten cycles, one-third of the atoms with the highest B factors were deleted from the model. In the initial cycles, the quality of the density was poor and the procedure mainly discarded ''bad'' atoms. The overall number of atoms in the model fell, the R factor went down, but R_{free} fell only slightly. Subsequently, when the model had improved by acquiring sufficient atoms in positions close to correct, the quality indicators improved more rapidly and after cycle 60 the procedure was stopped.

At this stage, the model contained about the number of atoms expected for the complete enzyme and associated solvent, and the R factor had fallen to 24.9 (35.6)%. Here the R factor is followed by R_{free} in parentheses. However, the *a posteriori* calculated average phase difference between this model and the final phase set was only 33.5°, whereas the average phase difference between the starting, molecular replacement and final sets was 80°. The resulting density map was of high quality and allowed easy building of the protein with almost all of its side-chains. This was done, as well as all other graphics display sessions, with the program QUANTA (Molecular Simulations Inc., San Diego, 1997).

Refinement

The unliganded and liganded enzyme models were refined with a similar protocol. The initial models obtained after molecular replacement (for the unliganded enzyme) or built into the uARP map (for the complex) were subjected to restrained refinement with REFMAC combined with automatic selection of waters by ARP. No σ cutoff was applied to the amplitudes. Electron density syntheses used were calculated on the basis of weighted observed and calculated structure factors, $(2mF_o - dF_c)$ or $(mF_o - dF_c)$. The protein restraints were taken from the library based on the Engh & Huber (1991) set.

The progress of the refinement was monitored by R and R_{free} (using 5% of all reflections) and by inspection of electron density maps. Some side-chain atoms were not visible in the maps and were given zero occupancy. They are mostly charged groups: Lys10 (CD, NE and CZ), Arg13 (NE, CZ, NH1 and NH2) and Met28 (CE) in the native model and Met1 (CE), Arg13 (NH1 and NH2), Lys42 (CE and NZ), Lys88 (NZ), Asp116 (CG, OD1 and OD2), Glu117 (CD, OE1 and OE2) and Asn118 (CG, OD1

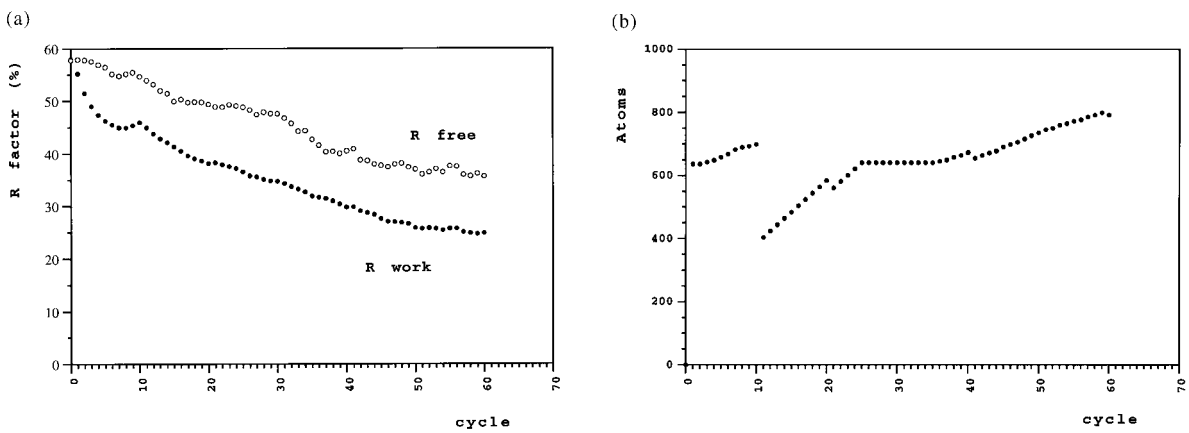

Figure 10. The progress of the unrestrained ARP. (a) The R and R_{free} factors and (b) the number of atoms in the model in successive cycles.

and ND2) in the complex. There was no interpretable electron density beyond residue Ile120 in the unliganded structure and beyond Asn118 in the complex. Some residues were built in double conformations with the affected atoms being assigned half occupancy: the side-chains of Ser32 and the Gly68-Gly69 peptide in the unliganded model and the side-chain of Cys22 and a main-chain region of three residues, Ser109, Asn110 and Ser110 in the complex.

The refinement converged and no new interpretable features were present in the electron density maps after 55 cycles of refinement of the native model, with $R_{cryst} = 16.10\,\%$ and $R_{free} = 21.55\,\%$. The refinement of the complex was stopped after cycle 60 with $R_{cryst} = 16.66\,\%$ and $R_{free} = 20.07\,\%$. At the end, five further refinement cycles (without updating the model by ARP) were performed against all data, giving a final R_{cryst} of 15.98 % for the unliganded and 16.86 % for the complexed structure.

Alignment and superposition of the known structures

All superpositions were performed using the program LSQKAB from the CCP4 suite (Collaborative Computattional Project, Number 4, 1994). The models of the unliganded and complexed EIAV dUTPase structures were superimposed using all their ordered CA atoms.

The three other known dUTPases were superimposed on the unliganded EIAV dUTPase in the following way. Initially they were overlapped using CA atoms of all residues corresponding with the sequence alignment, i.e. only the regions having evidently no counterparts in the other sequences were omitted. The overlaps were repeated iteratively discarding residues differing by more than 2 Å in the positions of their CA atoms.

Eventually, 53 structurally equivalent CA atoms were accepted, encompassing five different regions along the sequence corresponding to all or parts of secondary structural elements β2, βI, β3, βII, β4, β5B, β6 and β7, and turns 2 and 5 (Figure 3). There are only ten conserved residues among the 53. The residues used in the superpositions are shown in the Figure.

Protein Data Bank accession numbers

Atomic co-ordinates and structure factors have been deposited in the Brookhaven Protein Data Bank, entry code 1DUN and 1RDUNSF for unliganded EIAV dUTPase and 1DUC and 1RDUCSF for the complex with dUDP and Sr^{2+}. Codes for co-ordinates for other structures taken from PDB are: 1DUP, 1DUD, 1DUT. Co-ordinates for the human enzyme were provided by J. A. Tainer.

Acknowledgements

The authors gratefully thank Johan Nord for fruitful discussions. This research was supported by The Swedish Cancer Society, Magnus Bergwalls Stiftelse, Crafoords Stiftelse and Kungliga Fysiografiska Sällskapet i Lund. Partial support by the EC BIOMED contract BMH4-CT97-2711 is also acknowledged.

References

Barclay, B. J., Kunz, B. A., Little, J. G. & Haynes, R. H. (1982). Genetic and biochemical consequences of thymidylate stress. *Canad. J. Biochem.* **60**, 172-194.

Bergdoll, M., Remy, M.-H., Cagnon, C., Masson, J.-M. & Dumas, P. (1997). Proline-dependent oligomerization with arm exchange. *Structure*, **5**, 391-401.

Bergman, A.-C. (1997). PhD thesis, Lund University, Sweden.

Bergman, A.-C., Björnberg, O. , Nord, J., Rosengren, A. M. & Nyman, P. O. (1995). dUTPase from the retrovirus equine infectious anemia virus: high-level expression in *Escherichia coli* and purification. *Protein Expr. Purif.* **6**, 379-387.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. J., Brice, M. D., Rogers, J. K., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535-542.

Bertani, L. E., Häggmark, A. & Reichard, P. (1961). Synthesis of pyrimidine deoxyribonucleoside diphosphates with enzymes from *Escherichia coli*. *J. Biol. Chem.* **236**, P67-P68.

Bertani, L. E., Häggmark, A. & Reichard, P. (1963). Enzymatic synthesis of deoxyribonucleotides. *J. Biol. Chem.* **238**, 3407-3413.

Bessman, M. J., Lehman, I. R., Adler, J., Zimmerman, S. B., Simms, E. S. & Kornberg, A. (1958). Enzymatic synthesis of deoxyribonucleicacid. III. The incorporation of pyrimidine and purine analogues into deoxyribonucleicacid. *Proc. Natl Acad. Sci. USA*, **44**, 633-640.

Björnberg, O. & Nyman, P. O. (1996). The dUTPases from herpes simplex virus type 1 and mouse mammary tumour virus are less specific than the *Escherichia coli* enzyme. *J. Gen. Virol.* **77**, 3107-3111.

Broyles, S. S. (1993). Vaccinia virus encodes a functional dUTPase. *Virology*, **195**, 863-865.

Burley, S. K. & Petsko, G. A. (1985). Aromatic-aromatic interactions: a mechanism of protein structure stabilization. *Science*, **229**, 23-28.

Caradonna, S. J. & Adamkiewicz, D. M. (1984). Purification and properties of the deoxyuridine triphosphate nucleotidohydrolase enzyme derived from HeLa S3 cells. *J. Biol. Chem.* **259**, 5459-5464.

Cedergren-Zeppezauer, E. S., Larsson, G., Nyman, P. O., Dauter, Z. & Wilson, K. S. (1992). Crystal structure of a dUTPase. *Nature*, **355**, 740-743.

Cheevers, W. P. & McGuire, T. C. (1985). Equine infectious anaemia virus: immunopathogenesis and persistence. *Rev. Infect. Dis.* **7**, 83-88.

Collaborative Computational Project Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760-763.

Cruickshank, D. W. J. (1996). Protein precision re-examined: Luzzati plots do not estimate final errors. In *Proceedings of the CCP4 Study Weekend* (Dodson, E. J., Moore, M., Ralph, A. & Bailey, S., eds), pp. 11-22, SERC, Daresbury Laboratory, Warrington, UK.

Dauter, Z., Wilson, K. S., Larsson, G., Nyman, P. O. & Cedergren-Zeppezauer, E. S. (1998). The refined structure of dUTPase from *Escherichia coli*. *Acta Crystallog. sect. D*, **54**, 735-749.

Ducros, V., Brzozowski, A. M., Wilson, K. S., Brown, S. H., Ostergaard, P., Schneider, P., Yaver, D. S., Pedersen, A. H. & Davies, G. J. (1998). Crystal structure of the type-2 Cu depleted laccase from

Coprinus cinereus at 2.2 Å resolution. *Nature Struct. Biol.* **5**, 310-316.

Elder, J. H., Lerner, D. L., Hasselkus-Light, C. S., Fontenot, D. J., Hunter, E., Luciw, P. A., Montelaro, R. C. & Phillips, T. R. (1992). Distinct subsets of retroviruses encode dUTPase. *J. Virol.* **66**, 1791-1794.

El-Hajj, H. H., Zhang, H. & Weiss, B. (1988). Lethality of a *dut* (deoxyuridine triphosphatase) mutation in *Escherichia coli*. *J. Bacteriol.* **170**, 1069-1075.

Engh, R. A. & Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallog. sect. A*, **47**, 392-400.

Focher, F., Mazzarello, P., Verri, A., Hübscher, U. & Spadari, S. (1990). Activity profiles of enzymes that control the uracil incorporation into DNA during neuronal development. *Mutation Res.* **237**, 65-73.

Foil, L. D., Meek, C. L., Adams, W. V. & Issel, C. J. (1983). Mechanical transmission of equine infectious anemia virus by deer (*Chrysops flavidus*) and stable flies (*Stomoxys calcitrans*). *Amer. J. Vet. Res.* **44**, 155-156.

Gadsden, M. H., McIntosh, E. M., Game, J. C., Wilson, P. J. & Haynes, R. H. (1993). dUTP pyrophosphatase is an essential enzyme in *Saccharomyces cerevisiae*. *EMBO J.* **12**, 4425-4431.

Goulian, M., Bleile, B. M., Dickey, L. M., Grafstrom, R. H., Ingraham, H. A., Neynaber, S. A., Peterson, M. S. & Tseng, B. Y. (1986). Mechanism of thymineless death. *Advan. Exp. Med. Biol.* **195B**, 89-95.

Greenberg, G. R. & Somerville, R. L. (1962). Deoxyuridylatekinase activity and deoxyuridinetriphosphatase in *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **48**, 247-256.

Hagman, L.-O., Jansson, I. & Magnéli, C. (1968). The crystal structure of α-$Sr_2P_2O_7$. *Acta Chem. Scand.* **22**, 1419-1429.

Hochhauser, S. J. & Weiss, B. (1978). *Escherichia coli* mutants deficient in deoxyuridine triphosphatase. *J. Bacteriol.* **134**, 157-166.

Hoffman, I. (1987). PhD thesis, University of Saarbrücken, Germany.

Hokari, S., Sakagishi, Y. & Tsukada, K. (1982). Enhanced activity of deoxyuridine 5′-triphosphatase in regenerating rat liver. *Biochem. Biophys. Res. Commun.* **108**, 95-101.

Hunter, C. A. & Sanders, J. K. M. (1990). The nature of π-π interactions. *J. Amer. Chem. Soc.* **112**, 5525-5534.

Ingraham, H. A., Dickey, L. & Goulian, M. (1986). DNA fragmentation and cytotoxicity from increased cellular deoxyuridylate. *Biochemistry*, **25**, 3225-3230.

Kawakami, T., Sherman, L., Dahlberg, J., Gazit, A., Yaniv, A., Tronick, S. R. & Aaronson, S. A. (1987). Nucleotide sequence analysis of equine infectious anemia virus proviral DNA. *Virology*, **158**, 300-312.

Kornberg, A. & Baker, T. A. (1991). *DNA Replication*, 2nd edit., Freeman, New York.

Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.

Lamzin, V. S. & Wilson, K. S. (1997). Automated refinement for protein crystallography. *Methods Enzymol.* **277**, 269-305.

Larsson, G., Nyman, P. O. & Kvassman, J.-O. (1996a). Kinetic characterization of dUTPase from *Escherichia coli*. *J. Biol. Chem.* **271**, 24010-24016.

Larsson, G., Svensson, L. A. & Nyman, P. O. (1996b). Crystal structure of the *Escherichia coli* dUTPase in complex with a substrate analogue (dUDP). *Nature Struct. Biol.* **3**, 532-538.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the steroechemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283-291.

Lichtenstein, D. L., Rushlow, K. E., Cook, R. F., Raabe, M. L., Swardson, C. J., Kociba, G. J., Issel, C. J. & Montelaro, M. C. (1995). Replication *in vitro* and *in vivo* of an equine infectious anemia virus mutant deficient in dUTPase activity. *J. Virol.* **69**, 2881-2888.

Lindahl, T. (1982). DNA repair enzymes. *Annu. Rev. Biochem.* **51**, 61-87.

Matthews, B. W. (1968). Solvent content in protein crystals. *J. Mol. Biol.* **33**, 491-497.

McGeoch, D. J. (1990). Protein sequence comparisons show that the pseudoproteases encoded by poxviruses and certain retroviruses belong to the deoxyuridine triphosphatase family. *Nucl. Acids Res.* **18**, 4105-4110.

Mercer, A. A., Fraser, K. M., Stockwell, P. A. & Robinson, A. J. (1989). A homologue of the retroviral pseudoproteases in the parapoxvirus Orf virus. *Virology*, **172**, 665-668.

Mol, C. D., Harris, J. M., McIntosh, E. M. & Tainer, J. A. (1996). Human dUTP pyrophosphatase: uracil recognition by a β-hairpin and active sites formed by three separate subunits. *Structure*, **4**, 1077-1092.

Mosbaugh, D. W. (1988). Purification and characterization of porcine liver DNA polymerase. Utilization of dUTP and dTTP during *in vitro* DNA synthesis. *Nucl. Acids Res.* **16**, 5645-5658.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the Maximum Likelihood method. *Acta Crystallog. sect. D*, **53**, 240-255.

Navaza, J. & Saludijan, P. (1997). AMoRe: an automated molecular replacement program package. *Methods Enzymol.* **276**, 581-594.

Nord, J., Larsson, G., Kvassman, J. O., Rosengren, A. M. & Nyman, P. O. (1997). dUTPase from the retrovirus equine infectious anaemia virus: specificity, turnover and inhibition. *FEBS Letters*, **414**, 271-274.

Oldfield, T. J. (1992). SQUID: A program for the analysis and display of data from crystallography and molecular dynamics. *J. Mol. Graph.* **10**, 247-252.

Oroszlan, S., Rein, A. R. & Rice, N. R. (1991). The complete mapping of the *pol* gene of EIAV and the characterization of a new viral protein. In *Basic Research Program*, pp. 180-181, Annual Report, September 1991, National Cancer Institute, Frederick, MD .

Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.

Pardo, E. G. & Gutierrez, C. (1990). Cell cycle- and differentiation stage-dependent variation of dUTPase activity in higher plant cells. *Exp. Cell. Res.* **186**, 90-98.

Persson, R., Rosengren, A. M., Nyman, P. O., Dauter, Z. & Cedergren-Zeppezauer, E. S. (1997). Crystallization and preliminary X-ray crystallographic studies of dUTPase from equine infectious anemia virus. *Protein Pept. Letters*, **4**, 145-148.

Prasad, G. S., Stura, E. A., McRee, D. E., Laco, G. S., Hasselkus-Light, C., Elder, J. H. & Stout, C. D. (1996). Crystal structure of dUTP pyrophosphatase from feline immunodeficiency virus. *Protein Sci.* **5**, 2429-2437.

Preston, V. G. & Fisher, F. B. (1984). Identification of the herpes simplex virus type 1 gene encoding the dUTPase. *Virology*, **138**, 58-68.

Pri-Hadash, A., Hareven, D. & Lifschitz, E. (1992). A meristem-related gene from tomato encodes a dUTPase: analysis of expression in vegetative and floral meristems. *Plant Cell*, **4**, 149-159.

Ramakrishnan, C. & Ramachandran, G. N. (1965). Stereochemical criteria for polypeptide and protein chain conformation. *Biophys. J.* **5**, 909-933.

Richards, R. G., Sowers, L. C., Laszlo, J. & Sedwick, W. D. (1984). The occurrence and consequences of deoxyuridine in DNA. *Advan. Enzyme Reg.* **22**, 157-185.

Roberts, M. M., Copeland, T. D. & Oroszlan, S. (1991). *In situ* processing of a retroviral nucleocapsid protein by the viral proteinase. *Protein Eng.* **4**, 695-700.

Roseman, N. A., Evans, R. K., Mayer, E. L., Rossi, M. A. & Slabaugh, M. B. (1996). Purification and characterization of the vaccinia virus deoxyuridine triphosphatase expressed in *Escherichia coli*. *J. Biol. Chem.* **271**, 23506-23511.

Sancar, A. & Sancar, G. B. (1988). DNA repair enzymes. *Annu. Rev. Biochem.* **57**, 29-67.

Sevcik, J., Solovicova, A., Hostinova, E., Gasperik, J., Wilson, K. S. & Dauter, Z. (1998). The crystal structure of glucoamylase from *Saccharomycopsis fibuligera* at 1.7 Å resolution. *Acta crystallog. sect. D*, **54**, 854-866.

Shao, H., Robek, M. D., Threadgill, D. S., Mankowski, L. S., Cameron, C. E., Fuller, F. J. & Payne, S. L. (1997). Characterization and mutational studies of equine infectious anemia virus dUTPase. *Biochim. Biophys. Acta*, **1339**, 181-191.

Shlomai, J. & Kornberg, A. (1978). Deoxyuridine triphosphatase of *Escherichia coli*. *J. Biol. Chem.* **253**, 3305-3312.

Smith, H. G. (1953). The crystal structure of strontium hydroxide octahydrate, $Sr(OH)_2 8H_2O$. *Acta Crystallog.* **6**, 604-609.

Spector, R. & Boose, B. (1983). Development and regional distribution of deoxyuridine 5′-triphosphatase in rabbit brain. *J. Neurochem.* **41**, 1192-1195.

Steagall, W. K., Robek, M. D., Perry, S. T., Fuller, F. J. & Payne, S. L. (1995). Incorporation of uracil into viral DNA correlates with reduced replication of EIAV in macrophages. *Virology*, **210**, 302-313.

Stephens, R. M., Casey, J. W. & Rice, N. R. (1986). Equine infectious anemia virus *gag* and *pol* genes: relatedness to Visna and AIDS virus. *Science*, **231**, 589-594.

Strahler, J. R., Xhu, X.-X., Hora, N., Wang, Y. K., Andrews, P. C., Roseman, N. A., Neel, J. V., Turka, L. & Hanash, S. M. (1993). Maturation and stage proliferation-dependent expression of dUTPase in human T-cells. *Proc. Natl Acad. Sci. USA*, **90**, 4991-4995.

Threadgill, D. S., Steagall, W. K., Flaherety, M. T., Fuller, F. J., Perry, S. T., Rushlow, K. E., LeGrice, S. F. J. & Payne, S. L. (1993). Characterization of equine infectious anemia virus dUTPase: growth properties of a dUTPase deficient mutant. *J. Virol.* **67**, 2592-2600.

Tye, B.-K. & Lehman, I. R. (1977). Excision repair of uracil incorporated in DNA as a result of a defect in dUTPase. *J. Mol. Biol.* **117**, 293-306.

Tye, B.-K., Nyman, P. O., Lehman, I. R., Hochhauser, S. & Weiss, B. (1977). Transient accumulation of Okazaki fragments as a result of uracil incorporation into nascent DNA. *Proc. Natl Acad. Sci. USA*, **74**, 154-157.

Vertessy, B. G. (1997). The flexible glycine rich motif of *Escherichia coli* deoxyuridine triphosphate nucleotidohydrolase is important for functional but not for structural integrity of the enzyme. *Proteins: Struct. Funct. Genet.* **28**, 1-12.

Vertessy, B. G., Persson, R., Rosengren, A. M., Zeppezauer, M. & Nyman, P. O. (1996). Specific derivatization of the active site tyrosine in dUTPase perturbs ligand binding to the active site. *Biochem. Biophys. Res. Commun.* **219**, 294-300.

Vertessy, B. G., Larsson, G., Persson, T., Bergman, A.-C., Persson, R. & Nyman, P. O. (1998). The complete triphosphate moiety of nonhydrolyzable substrate analogues is required for a conformational shift of the flexible C-terminus in *Escherichia coli* dUTP pyrophosphatase. *FEBS Letters*, **421**, 83-88.

Wagaman, P. C., Hasselkus, L. C., Henson, M., Lerner, D. L., Phillips, T. R. & Elder, J. H. (1993). Molecular cloning and characterization of deoxyuridine triphosphatase from feline immunodeficiency virus (FIV). *Virology*, **196**, 451-457.

Williams, M. V. & Cheng, Y. (1979). Human deoxyuridine triphosphate nucleotidohydrolase. *J. Biol. Chem.* **254**, 2897-2901.

Williams, M. V., Holliday, J. & Glaser, R. (1985). Induction of a deoxyuridine triphosphate nucleotidohydrolase activity in Epstein-Barr virus-infected cells. *Virology*, **142**, 326-333.

Wilson, A. J. C. (1942). Determination of absolute from relative X-ray data intensities. *Nature*, **150**, 151-152.

Wohlrab, F. & Francke, B. (1980). Deoxyribopyrimidine triphosphatase activity specific for cells infected with herpes simplex virus type 1. *Proc. Natl Acad. Sci. USA*, **77**, 1872-1876.

Edited by K. Nagai

(Received 29 May 1998; received in revised form 12 October 1998; accepted 21 October 1998)